

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 03-014075

(43)Date of publication of application : 22.01.1991

(51)Int.Cl.

G06F 15/40

(21)Application number : 01-148438

(71)Applicant : RICOH CO LTD

(22)Date of filing : 13.06.1989

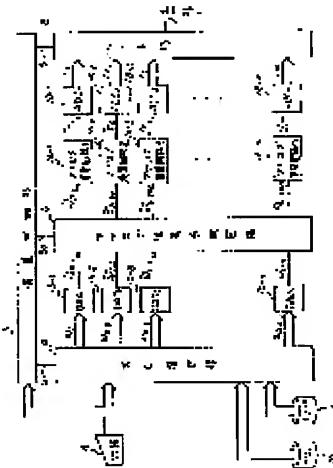
(72)Inventor : MORITA TETSUYA

(54) DOCUMENT RETRIEVAL DEVICE

(57)Abstract:

PURPOSE: To speed up the operation of the document retrieval device greatly by finding the evaluated value of each document by an analog arithmetic means by using key word relativity indicating the strength of relation between key words and the relation between key words and the document.

CONSTITUTION: This device is provided with a 1st storage means 10 which holds the strength of relation between key words, a 2nd storage means 12 which holds the relation between key words and documents, and a key word connection selecting means 4 which selects a key word connection according to retrieval conditions and converts it into an analog signal. Further, the device is provided with analog arithmetic means 16-1 – 16-m which calculate document likelihood by using the value of the key word connection in the form of the analog signal and converting means 18-1 – 18-m which convert the calculated document likelihood into a digital value. Consequently, the document likelihood is calculated by an analog circuit and fuzzy retrieval is performed fast.



⑯日本国特許庁(JP)

⑪特許出願公開

⑫公開特許公報(A)

平3-14075

⑬Int.Cl.⁵

G 06 F 15/40

識別記号

510 J

庁内整理番号

7313-5B

⑭公開 平成3年(1991)1月22日

審査請求 未請求 請求項の数 2 (全5頁)

⑮発明の名称 文書検索装置

⑯特 願 平1-148438

⑰出 願 平1(1989)6月13日

⑱発明者 森田 哲也 東京都大田区中馬込1丁目3番6号 株式会社リコー内

⑲出願人 株式会社リコー 東京都大田区中馬込1丁目3番6号

⑳代理人 弁理士 香取 孝雄 外1名

明細書

1. 発明の名称

文書検索装置

2. 特許請求の範囲

1. 条件式として指定されたキーワードに合う文書を、データベースから検索する文書検索装置において、該装置は、

前記キーワード間の関係の強さを示すキーワード間連度や該キーワードと前記文書間の関係を用いて各文書の評価値をアナログ演算手段により求めることを特徴とする文書検索装置。

2. 条件式として指定されたキーワードに合う文書を、データベースから検索する文書検索装置において、該装置は、

前記キーワード間の関係の強さを保持する第1の記憶手段と、

前記キーワードと前記文書の関係を保持する第2の記憶手段と、

検索条件に従ってキーワードコネクションを選択し、該選択したキーワードコネクションをアナ

ログ信号に変換するキーワードコネクション選択手段と、

前記アナログ信号に変換されたキーワードコネクションの値を用いて文書確度を計算するアナログ演算手段と、

該演算手段により計算された文書確度をデジタル値に変換する変換手段とを有することを特徴とする文書検索装置。

3. 発明の詳細な説明

【産業上の利用分野】

本発明は、利用者が指定した検索条件式に合致する文書を文書データベースから検索する際、検索結果を検索条件との合致の度合いの順に表示することにより適量・適質検索を行なうあいまい文書検索装置に関する。、

【従来の技術】

先に出願した出願番号63-185462に記載の「文書検索装置」は、キーワードコネクションと呼ばれるキーワード集を用いて連想検索やあいまい検

索を行なうものであった。また、ソフトウェア常識集IRシステムSOCKS(1)、(2)1988年、情報処理学会第35回全国大会予稿集第1505～1508頁に記載された従来技術では、ソフトウェア常識集をもとに、ソフトウェア開発における障害事例等から得られるキーワードを入力し、その原因に該当する教訓を連想的に検索するものであった。

これら従来技術の共通点はいずれも検索結果を「ある」「なし」の2値として表示するのではなく確度もしくは関連度と呼ばれる[0, 1]の値を結果として返すところに特徴があった。

[発明が解決しようとする課題]

しかし、このようなシステムを実現する場合、実数値の演算が多い上、計算量が文書数に比例するため汎用計算機には不向きである。

本発明はこの欠点を解消し、文書確度をアナログ回路を用いて計算し、高速な検索を行なうあいまい検索装置を提供することを目的とする。

[問題を解決するための手段]

本発明によれば、条件式として指定されたキー

を求める文書確度計算をアナログ演算回路で行なうことにより、検索速度を速くすることができる。

[実施例]

次に添付図面を参照して本発明による文書検索装置の実施例を詳細に説明する。

本発明による文書検索装置の実施例は、利用者からのキーワードを入力とし、検索結果として文書ごとの文書確度を計算しソートして利用者に提示する。文書確度は利用者が指定したキーワードとデータベース内にある文書との関係の強さを表わす評価値であり、先頃である出願番号63-185462に記載の「文書検索装置」において示された計算方式に基づいて与えられ、以下に示す式(1)の計算式のように表わせる。

$$r_i = \sum_{p \in Q \cap q \in D_i} w_{p,q} \quad \dots \quad (1)$$

r_i : 文書番号 i の文書確度

$w_{p,q}$: キー p とキー q の関係の強さ
(キーワード関連度)

Q : 検索用キーワードの集合

ワードに合う文書を、データベースから検索する文書検索装置は、キーワード間の関係の強さを示すキーワード関連度や該キーワードと文書間の関係を用いて各文書の評価値をアナログ演算手段により求める。

また本発明によれば、条件式として指定されたキーワードに合う文書を、データベースから検索する文書検索装置は、キーワード間の関係の強さを保持する第1の記憶手段と、キーワードと文書の関係を保持する第2の記憶手段と、検索条件に従ってキーワードコネクションを選択し、選択したキーワードコネクションをアナログ信号に変換するキーワードコネクション選択手段と、アナログ信号に変換されたキーワードコネクションの値を用いて文書確度を計算するアナログ演算手段と、演算手段により計算された文書確度をデジタル値に変換する変換手段とを有する。

[作用]

本発明によれば、あいまい文書検索装置において、多数のキーワード関連度を加算して文書確度

D_i : 文書番号 i の文書に含まれる
キーワードの集合

第1図には本実施例における文書検索装置の機能ブロック図が示されている。キーワード間の関係の強さを保持するキーワードコネクションファイル(KCF)10と、キーワードと文書の関係を保持する転置ファイル(IVTF)12は、キーワード間の関係の強さを示すキーワード関連度(KC)を選択するKC選択部4に接続されている。利用者が文書検索条件を入力する入力部14は、KC選択部4および演算制御部2に接続されている。演算制御部2は、KC選択部4、アナログ信号分配回路6およびソート部8にタイミング信号S1、S2およびS3をそれぞれ供給し、これらの回路間の入出力動作をすべて同期させる。なお、タイミング信号S2には接続制御信号も含まれる。DAコンバータ(DAC_{1~n})15はKC選択部4から入力した信号 $w_{p,i}$ をアナログ信号 $\tilde{w}_{p,i}$ に変換してアナログ信号分配回路6に出力する。アナログ演算回路16は信号分配回路6から入力されるアナログ信号 $\tilde{w}_{p,(t+i)}$

からアナログの文書確度 \tilde{r}_i を計算して出力する。ADコンバータ(ADC_{1~m}) 18は、文書確度をディジタル値 r_i に変換して出力する。これを入力するソート部 18は文書数分の文書確度を確度順にソートして出力する。次に本実施例の動作を説明する。

まず、利用者によって入力部から入力された検索用キーワードはKC選択部 4 に送られる。KC選択部 4 では、検索用キーワードの集合からキーワードを 1つ取り出しそのキーワードについて全てのキーワード間速度 KC を KCF10 から読み込みそれぞれ DAC 15 へ入力する。例えば、最初のキーワードが p であったとき $W_{p,1} \sim W_{p,k}$ (k はキーワード p についてのキーワード間速度の数) までが DAC 15-1 ~ 15-k へそれぞれ入力される。DAC 15 の数 n より KC の数が大きい場合すなわち $n < k$ のとき $W_{p,1} \sim W_{p,n}$ までが DAC 15-1 ~ 15-n へそれぞれ入力される。また、 $n > k$ の場合 $W_{p,k+1} \sim W_{p,n}$ として「0」が DAC 15-k+1 ~ DAC 15-n に入力される。

DAC 15 でアナログ信号に変換された $\tilde{W}_{p,1} \sim \tilde{W}_{p,n}$

と $\tilde{W}_{p,1} \sim \tilde{W}_{p,n}$ の総和が積分結果として得られる。しかしここで r_i は文書 i の文書確度であり、式 (1) で与えられたように $W_{p,q}$ のうち p が検索用キーワード集合に属しかつ q が文書 i に含まれるキーワード集合に属している間速度のみの総和をもとめる。キーワード p について全ての間速度のうち文書に含まれるキーワードによって $W_{p,q}$ の値を以下のように適宜変更することによって r_i は式 (1) のようになる。

$q \in D_i$ のとき、 $W_{p,q}$ を DAC に入力。

$q \notin D_i$ のとき、 0 を DAC に入力。

以上のような制御を接続制御信号と同期して行なうことにより n 時間単位でキーワード p について適切な $W_{p,q}$ を積分できる。さらに検索用キーワード集合に含まれるキーワード数 * n 時間単位で m 文書確度をすべて計算できる。

たとえば入力部 14より K 個の検索用キーワードからなる集合 Q が入力された場合を考える。

文書確度 r_i は式 (1) に示したように与えられる。KC選択部 4 では、まず集合 Q から 1つのキー

は、アナログ信号分配回路 6 に入力される。第 2 図は、アナログ信号分配回路 6 の 1 接続例である。これは DAC 15 の数 n に対し、アナログ演算回路 16 の数 m が少ない場合の構成例となっている。ここで記号 \oplus は $(\text{modulo } n \text{ の加算結果} + 1)$ の値を返す演算子とする。アナログ信号分配回路 6 は、接続制御信号によって接続状態を変える。例えば、 $t = 1$ のとき入力 $\tilde{W}_{p,1} \sim \tilde{W}_{p,m}$ がそのまま出力 1 ~ m に接続され、 $t = 2$ のとき入力 $\tilde{W}_{p,2} \sim \tilde{W}_{p,m+1}$ が出力 1 ~ m に接続される。第 2 図の例は、 $t = n-3$ の接続状態を示している。つまり所定の時間単位ごとに接続制御信号 S2 を入力し、これを n 回繰り返すことにより、出力 1 ~ m に $\tilde{W}_{p,1} \sim \tilde{W}_{p,n}$ をそれぞれ 1 時間単位ずつ接続することができる。

第 3 図は、アナログ演算回路 16 の構成例であり反転出力のアナログ積分回路となっている。入力 $\tilde{W}_{p,t}$ は前述の時間単位ずつ積分され積分結果は、 $-\tilde{r}_i$ として出力される。前述のように $\tilde{W}_{p,1} \sim \tilde{W}_{p,n}$ が入力されている場合、時間単位を n 回繰り返す

ワードを選択する。今これをキーワード p としたとき、KC選択部 4 から DAC 15 に入力される値は時間単位 t によって以下のように変化する。

DAC15-1 : $(t + 0) \in D_{t+0}$ のとき、 $W_{p,t+0}$ を DAC 15-1 に入力。

$(t + 0) \notin D_{t+0}$ でないとき、 0 を DAC 15-1 に入力。

DAC 15-2 : $(t + 1) \in D_{t+1}$ のとき、 $W_{p,t+1}$ を DAC 15-2 に入力。

$(t + 1) \notin D_{t+1}$ でないとき、 0 を DAC 15-2 に入力。

⋮
⋮
⋮

DAC 15-i : $(t + i - 1) \in D_{t+i-1}$ のとき、 $W_{p,t+i-1}$ を DAC 15-i に入力。

$(t + i - 1) \notin D_{t+i-1}$ でないとき、 0 を DAC 15-i に入力。

ただし、 $(t + i - 1)$ はキーワード $(t + i - 1)$ を表わす。

以上の処理を検索用キーワード Q に含まれるす

べてのキーワードについて、それぞれ n 時間単位繰り返すことにより式(1)で示した文書確度が m 個の文書について計算できる。

[発明の効果]

本発明によれば、汎用計算機で浮動少数演算を行なう場合と比べ、アナログ演算回路を使用することにより大幅な高速化が計れる。特に本発明の場合、高精度は要求されず 8 ~ 12 ビット程度のアナログ演算精度で十分である。また、並列化およびパイプライン化が容易であり文書データの増大にも対応が可能である。

4. 図面の簡単な説明

第1図は本発明のあいまい文書検索装置の実施例を示す機能ブロック図。

第2図は、第1図におけるアナログ分配回路の説明図。

第3図は、第1図のアナログ演算回路の構成例を示す回路図である。

主要部分の符号の説明

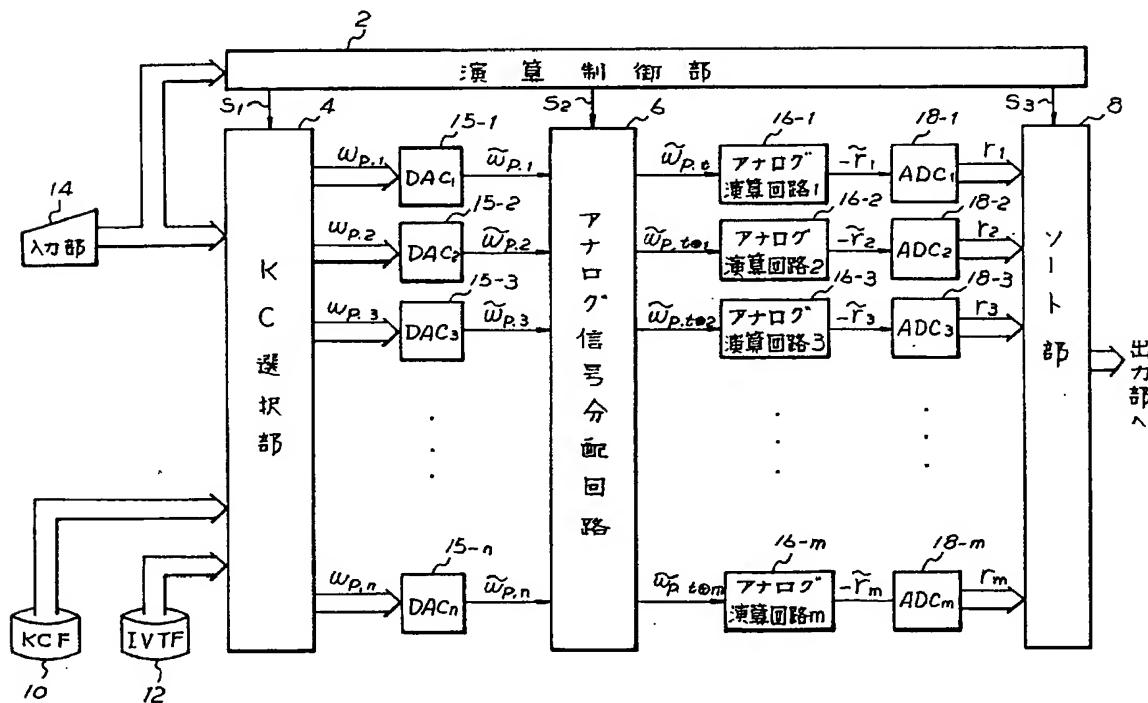
2. . . 演算制御回路

- 4. . . KC選択部
- 6. . . アナログ信号分配回路
- 8. . . ソート部
- 10. . . キーワード接続ファイル
- 12. . . 転置ファイル
- 14. . . 入力部
- 15. . . DAコンバータ
- 16. . . アナログ演算回路
- 18. . . ADコンバータ

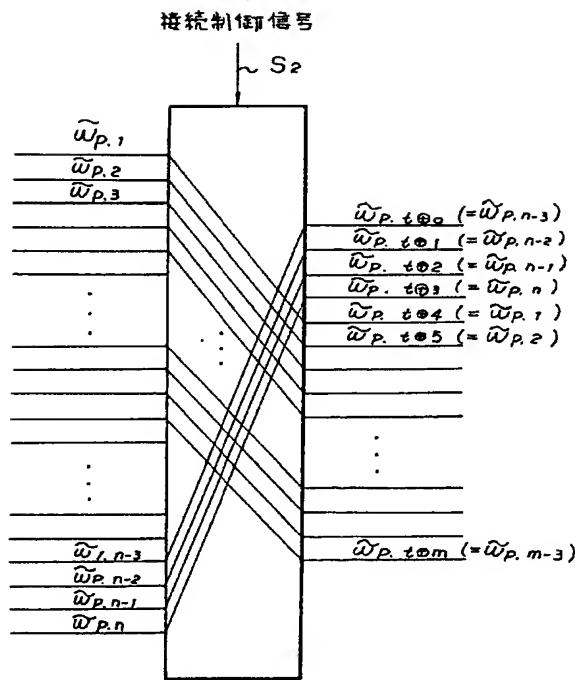
特許出願人 株式会社リコー

代理人 香取 幸雄
丸山 隆夫

第1図



第2図



第3図

